DOCUMENT RESUME

ED 235 802 IR 050 458

AUTHOR McGrath, William E.; Hickey, Thomas B.
TITLE Multidimensional Mapping of Libraries Based on Shared Holdings in the OCLC Online Union Catalog. Research Report.
INSTITUTION OCLC Online Computer Library Center, Inc., Dublin, Ohio.
REPORT NO OCLC/OPR/RR-83/5
PUB DATE 1 Jul 83
NOTE 52p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cluster Analysis; *Databases; Library Acquisition; *Library Collections; *Library Networks; *Multidimensional Scaling; Online Systems; Union Catalogs
IDENTIFIERS Bibliographic Utilities; *OCLC; *Shared Resources

ABSTRACT

This report describes the multidimensional scaling (MDS) method for interpreting collection overlap based on the number of books held by pairs of libraries in large groups (as many as 100 libraries or more). MDS results for four groups of libraries sampled from the OCLC database are also presented. MDS is described as a class of data-analytic or descriptive techniques designed to uncover the structure of a database in terms of clusters of similar objects within one or more dimensions. It is reported that (1) in the first group of libraries studied (a random sample of 60 network libraries), public and academic libraries separated in two dimensions; (2) in a random sample of 50 academic libraries, types of libraries (law, medical, college, university, and denominational) separated in two dimensions; (3) a random sample of 100 libraries without regard to type, nevertheless clustered in two dimensions by library type (university, college, public, denominational, medical, and special); and (4) a random sample of 50 public libraries had tentative clusters not yet identified. The implications of these findings are discussed for cooperative collection development, interlibrary lending, partitioning of the OCLC database by type of library, and library network relationships. A 22-item bibliography and lists of the libraries included in the study are also provided. (Author/ESR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED235802

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Report Number: OCLC/OPR/RR-83/5
Date: 1983 July 1

Research Report Prepared for OCLC

on

Multidimensional Mapping of Libraries Based on Shared Holdings in the OCLC Online Union Catalog

by

William E. McGrath
State University of New York at Buffalo

and

Thomas B. Hickey
Office of Research

IR050458

OCLC Online Computer Library Center, Inc.
Office of Research
6565 Frantz Road
Dublin, OH 43017-0702

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Cray Little
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

OCLC is a not-for-profit computer library service and research organization based in Dublin, Ohio. The Center operates an international network that libraries use to acquire and catalog books, order custom-printed catalog cards, arrange interlibrary loans, and maintain location information on library materials.

All rights reserved; however, individuals may quote from or reproduce portions of the text for noncommercial, educational, or private study or research. Full credit must be given to both the author(s) and the Office of Research, OCLC, Dublin, Ohio.

RESEARCH REPORT SERIES

Through the Research Report Series, OCLC research project results are made available to the public. OCLC Research Reports are published and distributed by OCLC. In addition, the reports are deposited with ERIC, which makes them available in microfiche as well as paper copy. Until Research Reports are available from ERIC (usually six to nine months after publication), OCLC will supply copies to persons and institutions requesting them. Due to the increasing demand for these reports and increasing costs to produce them, OCLC charges requestors a nominal fee. The OCLC price or ERIC number is noted below for each report.

Subject Heading Patterns in OCLC Monographic Records
by E. O'Neill and R. Aluri
Report Number: OCLC/RDD/RR-79/1; ERIC ED 183167; 96 pages

An Overview of a Proposed Monitoring Facility for the Large-scale, Network-based OCLC Online System
by W. Dominick, D. Penniman, and J. Rush
Report Number: OCLC/OPR/RR-80/1; ERIC ED 186042; 30 pages

Analytical Review of Catalog Use Studies
by K. Markey
Report Number: OCLC/OPR/RR-80/2; ERIC ED 186041; 66 pages

A Method for Correcting Typographical Errors in Subject Headings in OCLC Records
by E. O'Neill and R. Aluri
Report Number: OCLC/OPR/RR-80/3; ERIC ED 215680; 34 pages

Field, Subfield, and Indicator Statistics in OCLC Bibliographic Records
by T. B. Hickey
Report Number: OCLC/OPR/RR-81/1; ERIC ED 202484; 156 pages

Development of a Probabilistic Author Search and Matching Technique for Retrieval and Creation of Bibliographic Records
by T. B. Hickey
Report Number: OCLC/OPR/RR-81/2; ERIC ED 206286; 48 pages

Survey of Online Searching Instruction in Schools of Library and Information Science
by C. H. Fenichel and S. P. Harter
Report Number: OCLC/DD/RR-81/3; ERIC ED 210038; 108 pages

The American Public Library
by D. Sager
Report Number: OCLC/OPR/RR-82/1; ERIC ED 221222; 58 pages

The Process of Subject Searching in the Library Catalog: Final Report of the Subject Access Research Project
by K. Markey
Report Number: OCLC/OPR/RR-83/1
OCLC Price: $7.00; 152 pages

Current Utilization of Online Catalogs: Transaction Log Analysis; Final Report to the Council on Library Resources, Volume I
by J. E. Tolle
Report Number: OCLC/OPR/RR-83/2
OCLC Price: $6.50; 128 pages

Online Catalog Use: Results of Surveys and Focus Group Interviews in Several Libraries;
Final Report to the Council on Library Resources, Volume II
by K. Markey
Report Number: OCLC/OPR/RR-83/3
OCLC Price: $12.50; 284 pages

A Comprehensive Study of Online Public Access Catalogs:
An Overview and Application of Findings;
Final Report to the Council on Library Resources, Volume III
by N. K. Kaske and N. P. Sanders
Report Number: OCLC/OPR/RR-83/4
OCLC Price: $5.50; 102 pages

Multidimensional Mapping of Libraries Based on Shared Holdings in the OCLC Online Union Catalog
by W. E. McGrath and T. B. Hickey
Report Number: OCLC/OPR/RR-83/5
OCLC Price: $3.50; 54 pages

OCLC LIBRARY, INFORMATION, AND COMPUTER SCIENCE SERIES

The OCLC Library, Information, and Computer Science Series of monographs is another vehicle through which the OCLC Office of Research communicates the results of investigations and discusses issues. The first monograph was published in June 1982; the second in June 1983. The third monograph, an analysis of the number of terminals needed to support an online public access catalog, will be available in fall 1983. Additional monographs are planned for publication in 1983-84.

[1] Online Public Access Catalogs: The User Interface
by C. R. Hildreth
Product Number: 801
Price: $18.00; 280 pages; softbound

2. Public Library Administrators' Planning Guide to Automation
by D. J. Sager
Product Number: 802
Price: $12.50; 144 pages; softbound

ORDERING INFORMATION

Orders for OCLC publications must be prepaid, unless to be shipped and billed to an OCLC member.
Prepaid orders--send check and order to OCLC, DEPT BOX ONB, COLUMBUS, OHIO 43265.
Orders to be billed--send order, with your OCLC symbol and network indicated, to OCLC DOCUMENTATION DEPT, 6565 FRANTZ ROAD, DUBLIN, OH 43017-0702.

OCLC publications with ERIC numbers assigned should be ordered from ERIC DOCUMENT REPRODUCTION SERVICE, PO BOX 190, ARLINGTON, VA 22210.

ACKNOWLEDGMENTS

This research was conducted while the principal author was the OCLC Visiting Distinguished Scholar during 1979-80. Grateful acknowledgment is given to OCLC staff for their considerable support and to the organization as a whole for its generosity in making its facilities available, including access to data in the Online Union Catalog. The author wishes to thank particularly Dr. Neal Kaske, Director, Office of Research, without whom the work could not have been done. The author also thanks Professors Thomas Nygren and Robert MacCallum of The Ohio State University's Department of Psychology for their assistance and the author's use of their computer programs. Acknowledgment is also given to OCLC Research Assistants, Chi-shen Jane Kohn, William Stalcup, and Pat Hartsel. Finally, the ALSCAL program written by Forrest Young and Rostyslaw Lewyckyj of the University of North Carolina was crucial in computing the multidimensional configurations.

ABSTRACT

How many books does a group of libraries have in common? In early and pre-OCLC days, the question was simple and the answer vital. If a large number were held in common, there would be a compelling argument for computerized bibliographic databases for shared cataloging and interlibrary loan. "Overlap" studies did provide argument, and the rest is history.

Now, with more than 9 million records in the OCLC database and more than 3,000 member libraries using OCLC's Online Library Systems, the question, while far more complex, is still of great interest. With substantial "overlap" an established fact, how could a group of libraries know if they were acquiring books at a rate which optimally met the joint needs of their clientele? Would the nature of this overlap warrant establishing cooperative collection development? Should libraries change their interlibrary loan policy? Could they enhance their retrieval procedures in any way?

In the early studies, answers were usually given in terms of percentages, e.g., 85% of library A's books were also held by library B. In a small group of libraries, the number of these percentages (that is, the number of pairs of libraries) was small and, therefore, manageable. For 10 libraries, the number of pairs is 45. But for 50 libraries, the number of pairs is 1225, far too many to interpret separately.

This report describes a method--multidimensional scaling (MDS)--for interpreting overlap based on the number of books held by pairs of libraries in large groups (as many as 100 libraries or more). MDS results for four groups of libraries sampled from the OCLC database are also presented. In the first group (60 libraries in the AMIGOS and NELINET networks), public and academic libraries clearly separated in two dimensions. In the second (a random sample of 50 academic libraries), types of libraries--law, medical, college, university, and denominational--clearly separated in two dimensions. In a third group, a random sample of 100 libraries without regard to type, the libraries nevertheless clustered in two dimensions by university, college, public, denominational, medical, and special. A fourth group, a random sample of 50 public libraries, had tentative clusters not yet identified.

These results could be used as the basis for further studies in subject access, cooperative collection development, interlibrary loan, partitioning of the database by type of library, and in library network relationships.

CONTENTS

ACKNOWLEDGMENTS, vi

ABSTRACT, vii

ILLUSTRATIONS, xi

TABLES, xi

1.0 INTRODUCTION, 1

2.0 BACKGROUND AND PURPOSE, 2

3.0 METHODOLOGY, 4

- 3.1 Limitations of Early Overlap Studies, 4
- 3.2 Multidimensional Scaling, 4
- 3.3 Data, 6
  - 3.3.1 The 1% Sample, 6
  - 3.3.2 Subsamples of Libraries, 6
  - 3.3.3 Data Formatting (Symmetric Matrices), 6
- 3.4 Data Analysis, 7

4.0 RESULTS, 9

- 4.1 60 Network Libraries, 9
- 4.2 50 Academic Libraries, 12
- 4.3 50 Public Libraries, 18
- 4.4 100 Randomly Selected Libraries, 20

5.0 IMPLICATIONS, APPLICATIONS, and CONCLUSIONS, 25

- 5.1 Network Libraries, 25
- 5.2 Academic Libraries, 25
- 5.3 Public Libraries, 27
- 5.4 100 Randomly Selected Libraries, 27
- 5.5 Contribution of Multidimensional Scaling and Other Techniques, 27

6.0 FURTHER RESEARCH, 30

REFERENCES, 31

APPENDIX A. Random Sample List of 60 Network Libraries, 33

APPENDIX B. Random Sample List of 50 Academic Libraries, 35

APPENDIX C. Random Sample List of 50 Public Libraries, 37

APPENDIX D. Random Sample List of 100 Libraries, 39

ILLUSTRATIONS

1. STRESS and R-Square Values for 60 Network Libraries, 10
2. Two-dimensional Configuration of 60 Libraries in Two Networks, 11
3. STRESS and R-Square Values for 50 Academic Libraries, 13
4. Generalized Venn Diagram in Early Overlap Studies, 14
5. Two-dimensional Configuration of 50 Academic Libraries, 15
6. STRESS and R-Square Values for 50 Public Libraries, 18
7. Two-dimensional Configuration of 50 Public Libraries, 19
8. STRESS and R-Square Values for 100 Randomly Selected Libraries, 22
9. Two-dimensional Configuration of 100 Randomly Selected Libraries, 23

TABLES

1. Combinations of Libraries, 4
2. Options Used in the ALSCAL Program, 7
3. STRESS and R-Square Values for 60 Network Libraries, 9
4. STRESS and R-Square Values for 50 Academic Libraries, 12
5. Graphic Distribution of 50 Academic Libraries, 17
6. STRESS and R-Square Values for 50 Public Libraries, 18
7. Graphic Distribution of 50 Public Libraries, 20
8. STRESS and R-Square Values for 100 Randomly Selected Libraries, 21

Report Number: OCLC/OPR/RR-83/5
Date: 1983 July 1

1.0 INTRODUCTION

This report examines both the similarities and the dissimilarities of holdings in the OCLC Online Union Catalog (also referred to as the OCLC bibliographic database) that are common to each library in four samples of OCLC member libraries. Patterns of these similarities and dissimilarities can be described in terms of spatial configurations; that is, they can be described in terms of computed "distances" and then displayed on "maps." The distances are mathematical, not geographic, and the maps are plotted in one, two, three, and four dimensions.

The results--clusters of libraries and conceptual dimensions with Cartesian coordinates--should be useful in determining optimum holdings for shared cataloging and interlibrary loan, with implications for cooperative collection development, cost sharing, network organization by type of library, structuring of databases (to reduce retrieval time), and theory building.

2.0 BACKGROUND AND PURPOSE

When the library profession began thinking seriously about forming cooperative networks in the fifties and sixties, some basic questions about common holdings had to be answered. An important realization was that if cataloging information were available to every library, large cost and time savings could be realized (Kilgour 1969). But librarians first wanted to know the extent to which they were duplicating one another's efforts. For example, for a group of libraries, they wanted to know how many owned the same title, and how many titles were duplicated in the group. This kind of information was needed for cost studies on which to base decisions about the feasibility of shared cataloging.

These early studies were often called "overlap studies." Merritt (1942) was one of the first to hypothesize overlap relationships. Nugent's study (1968) of the holdings common to the six New England state university libraries provided information vital to the establishment of NELINET, Inc. Other studies included those conducted by Leonard, Maier, and Dougherty for Colorado libraries (1969), McGrath, Simon, and Dickinson for Louisiana libraries (1971), Altman for secondary school libraries (1972), Davis and Shaw for Canadian and U.S. libraries (1979), and Evans, Beilby, and Gifford for SUNY libraries (1978).

Typically, these studies computed the average number of copies per title, the percent of titles held in common by 1, 2, 3, and n libraries, and the average overlap in the entire system. All established, to one degree or another, that individual libraries did indeed enjoy a certain amount of uniqueness and that a predictable portion (though sometimes with contradictory interpretation) of their holdings was similar. For example, a 50% overlap in a system might be considered "a tremendous amount" by some authorities, but its reciprocal, 2 (defined as an average of two copies per title), might be considered by others as "extremely low."

With today's cooperative networks firmly established and large amounts of holdings data readily available online, we should like to know what new insights may be gleaned from a closer analysis, using up-to-date statistical methods, to answer some of the more important questions.

For example, the more an online record is consulted for cataloging, the greater the economic benefit and the lower the unit cost of that record. This implies that the more libraries having a copy of that title, the better. On the other hand, the economics of interlibrary lending indicate that not every library needs to own the book, and thus a smaller number of locations would be better. These indications suggest that some optimum number of locations per title would best serve both cataloging and interlibrary loan.

Holdings overlap, or holdings similarity or dissimilarity, also has implications for cooperative collection development. For example, a group of very similar libraries (large overlap) may opt to diversify their holdings according to some agreed-upon scheme. Or a group of homogeneous libraries (for example, medical) may wish to pool their budgets to build greater depth in subject areas of common interest.

Libraries in a group may also wish to learn more about characteristics (e.g., academic program, type of patronage, and community profiles) that make their holdings similar or dissimilar. This knowledge may help them to make better policy decisions and to make their budgets go further.

This report does not answer these questions in any great detail; instead, it describes procedures for the benefit of those who might wish to study these or similar questions. The methodology for examining similarities and dissimilarities is presented, and sample results, some implications, and a few conclusions are given.

3.0 METHODOLOGY

3.1 Limitations of Early Overlap Studies

In early overlap studies, a key analytical statistic was the "percentage" (the percentage of overlap overall, or the percentage between two or three libraries, for example). If the number of libraries was very small (fewer than ten), this statistic and its methodology were adequate and their interpretation meaningful. For a larger number of libraries, however, the large number of percentages was not readily interpretable. This shortcoming can be shown mathematically (see table 1), from the formula for combinations,

$$C(n,m) = \frac{n}{m!(n-m)!}$$

where C(n,m) is the number of combinations of n things taken m at a time. Clearly, when percentage combinations are treated as results, any number of libraries over ten yields uninterpretable masses of such percentages.

3.2 Multidimensional Scaling

If these "results" are treated as data input for further analysis, however, then some comprehensibility can be extracted. One method currently popular

TABLE 1

Combinations of Libraries

Number of Libraries (n)	Combinations of Two (m=2)	Combinations of Three (m=3)	Total Combinations (m)
6 (Nugent study)	15	20	64
10 (Evans study)	45	120	1024
50* (Small network)	1225	19,600	billions
60* (Small network)	1770	34,220	billions
100* (Typical network)	4950	161,700	astronomical

*These are actual sample sizes used in this study.

with psychologists, sociologists, market analysts, and others is multidimensional scaling (MDS).

MDS is a class of data-analytic or descriptive techniques designed to uncover the hidden structures of databases--that is, to yield a configuration of objects in space. MDS uses data described variously as "proximities," "similarities," or "dissimilarities" between any group of objects (e.g., persons, cars, foods, or cities) (Torgerson 1958). Data can be the number of things held in common by two persons, correlation coefficients, percentages, or any dyadic measure. Input data are invariably displayed in a matrix. Matrices can be of several sorts, but the most common, and the one used in this report, is a simple, two-way, (n x (n-1))/2 matrix, which is, in fact, the very type of symmetric matrix used to display output from the early overlap studies.

MDS is related to factor analysis, although the mathematical techniques are quite different. Factor analysis is parametric. As such, it is subject to the usual restrictions of homogeneity of variance, the requirement of a test statistic, and so on. On the other hand, MDS is nonparametric, distribution- free, and far more flexible. Classical MDS, as developed by Torgerson, uses the least squares technique and interval or ratio data to compute Euclidean distances between objects. More recently, powerful techniques have been developed by Shepard (1962) and Kruskal (1964) that use ordinal or nonmetric (i.e., noninterval, nonratio) data to compute distances between objects. Elements of both the metric and nonmetric techniques have been combined in a very flexible computer program called ALSCAL, written by Young and Lewyckyj, the program used in this study (1979).

The structure of the database uncovered by MDS analysis is described in terms of clusters and dimensions. Similar objects will cluster together in space; dissimilar objects will be distant from each other. Furthermore, the objects will align themselves along one or more dimensions orthogonal to each other. Sometimes, the characteristic that binds or separates the objects from each other in each dimension can be identified.

One way to understand the intent of MDS is to compare it to a road atlas' table of miles between cities. Using miles as input data, MDS can recover and, if necessary, draw the map, even if the statistician has no preconception of the map. MDS can generate the map in one dimension (e.g., east-west), showing the locations of all the cities on a single line; in two dimensions (e.g., east-west and north-south); and in three dimensions (e.g., east-west, north-south, and altitude). Mathematically, four or more dimensions can be generated; however, this dimensional multitude is rarely needed. This hypothetical map would also show clusters of cities, such as those in the northeast corridor.

MDS uses a test statistic called STRESS. It is a measure of fit (between the original data--similarities or dissimilarities--and the computed distances) for each dimensional solution. Since it is a least squares procedure, the smaller the STRESS the better. The higher dimensional solutions necessarily yield smaller STRESS. Acceptance of a STRESS value is largely a judgmental one. Criteria for acceptance are given by Kruskal and Wish (1978). ALSCAL uses a variation of STRESS called SSTRESS (spelled with two S's) to indicate improvement of fit in each iteration, but the

final fit is given by STRESS. ALSCAL also computes and recommends the more familiar RSQ, or squared multiple correlation coefficient.

3.3 Data

3.3.1 The 1% Sample

The sample used in this study consists of a 1% systematic sample of bibliographic records and their associated holdings records from the OCLC Online Union Catalog as of 1980 January 2. The sample was obtained by extracting all records with an OCLC control number ending in "44." The sample, therefore, contained records with control numbers 44, 144, 244, etc., with the exception of any deleted records. There were 58,375 bibliographic records in the sample.

3.3.2 Subsamples of Libraries

From the 1979 list of 2,300 OCLC member institutions, four samples of libraries were taken. These libraries are listed in the Appendixes. Sample 1 contains 15 public and 15 academic libraries randomly selected from each of two geographically distant OCLC member networks, AMIGOS in the Southwest and NELINET in New England, for a total of 60 libraries. Sample 2 contains 50 randomly selected academic libraries. Sample 3 contains 50 randomly selected public libraries, while Sample 4 comprises 100 libraries randomly selected without respect to type or geographic location.

3.3.3 Data Formatting (Symmetric Matrices)

The programs for scanning the sample database and formatting it for transmission to the ALSCAL package were written in PASCAL, FORTH, and APL for execution on a Xerox Sigma 9 computer. The first task was to extract the holdings records from the sample.

Every bibliographic record that is known to be held by one or more institutions has a holdings record associated with it. This record contains a compressed array of bits; each bit position corresponds to OCLC's institution number assigned to cataloging participants. After the holdings records, which were intermixed with bibliographic records in the sample, were separated, their compressed bit arrays were decompressed to simplify subsequent processing and stored on disk.

A program was then written in the FORTH language that could be initialized with a list of institution numbers for which counts of the number of commonly held records were to be made. This program, as well as the decompression program that preceded it, was written in FORTH because of that language's nearly unique ability to implement efficiently a variety of data structures. This program's counts of co-occurrence were kept in a triangular array corresponding to the upper half of the symmetrical square matrix, and the holdings records were accessed via a bit-addressable array.

After all the holdings records had been scanned, the triangular array was written out to disk. This array was then read in by an APL program that

could do some data manipulation, such as eliminating rows/columns with less than some threshold of counts. (This was desirable because libraries with very small shared holdings had a tendency, because of sampling error, to be positioned far from other libraries. This positioning forced all other libraries to cluster densely near the center of the resulting plot. The algorithm, although not used in the ALSCAL analysis reported in this document, should be used in future analyses.) Finally, the data were written out in a format suitable for processing by the statistical package used for data analysis.

3.4 Data Analysis

The ALSCAL program used in this study may be used for virtually any type of MDS, and program options are specified by the user. An explanation of some of the options is given in table 2. Those not discussed are self-explanatory. The general model is Euclidean, and the matrix is a single (n x (n-1))/2 symmetric matrix of similarities (Young and Lewyckyj 1979).

TABLE 2

Options Used in the ALSCAL Program

Program Option	Description
Number of Row and Column Stimuli:	"Stimuli" mean "objects." (Here, stimuli are the number of libraries. Rows and columns are the same.)
Number of Matrices:	One for each sample of libraries.
Measurement Level:	Ratio (i.e., the ratio property of the original data was used, as opposed to interval, ordinal, or nominal).
Data Type:	Symmetric-similarity, upper half of the matrix is redundant. The greater the number of books held in common by two libraries, the more similar the libraries and the closer on the maps.
Measurement Process:	Holdings data could be regarded as either discrete or continuous, but the continuous option unties tied data, resulting in better fit when ordinal data are used; not really relevant here.
Measurement Conditionality:	Other options (row conditional and unconditional) not appropriate in this model; though matrix conditional is the default, it is relevant only when there is more than one matrix.

TABLE 2--Continued

Program Option	Description
Model Type:	Simple Euclidean using a two-way matrix.
Maximum Number of Dimensions:	Although specifying more than four dimensions would undoubtedly yield a better fit, it would also yield a very large printout and would be more difficult to interpret; for satisfactory interpretation (with 50 and 60 objects), more than four are probably not necessary).
Minimum Number of Dimensions:	No fewer than one dimension can be specified; with 50 or 60 objects, two is probably the minimum for satisfactory interpretation.
Negative Weights:	Weights, whether negative or nonnegative, do not apply to the Euclidean model used in this report; a nonnegative restriction is the default.
Maximum Number of Iterations:	Relatively few iterations (probably under 10) are needed in the Euclidean metric model.
Convergence Criterion:	Refers to the improvement of SSTRESS. When improvement is less than the amount specified (in this report: 0.001), computation halts; 0.01 is probably sufficient.
I/O Options:	Print, plot, etc., are self-evident.
Initial Stimulus Coordinates:	Input coordinates needed only if known coordinates might be improved or if ordinal data would result in too many iterations; rows and columns are the same.
Initial Column Coordinates:	This instruction is ignored for the symmetric model.
Initial Subject Weights:	Does not apply to the model used here.
Initial Stimulus Weights:	Does not apply to the model used here.

All MDS runs were done on The Ohio State University's IBM 370 computer using JCL input instructions. The ALSCAL program was supplied by Professor Robert MacCallum of OSU's Department of Psychology.

4.0 RESULTS

Coordinates were computed for one, two, three, and four dimensions. These coordinates and their printed configurations may be obtained from the authors.

A summary of results for the four samples, with some analysis of the samples for the 60 network libraries and 50 academic libraries, is given in this section.

The scope of the samples must be carefully qualified. Results reflect only the similarities of holdings in the OCLC Online Union Catalog. Since some of the libraries have holdings that are not in this catalog, it cannot be readily inferred that similarities based on total holdings are like those found here.

4.1 60 Network Libraries

Table 3 shows STRESS and R-Square for the 60 network libraries' sample. Figure 1 shows that each higher-dimensional solution yields improvement in both STRESS and R-Square, but the greatest improvement occurs with the two-dimensional solution, shown by the sharper angle between dimensions one and three. One dimension is inadequate to explain overlap relationships among the 60 libraries.

TABLE 3

Configurations of STRESS and R-Square Values for 60 Network Libraries

Configuration (Dimension)	STRESS	R-Square	Comment
Four	0.17	0.88	Excellent, but probably not interpretable
Three	0.22	0.83	Good
Two	0.29	0.77	Adequate and interpretable
One	0.45	0.59	Inadequate

The slight improvement with three and four dimensions (R-square improvement = 0.05 and 0.06 respectively) suggests that no more than two are needed for a satisfactory explanation of the holdings relationships. Reliance on STRESS and R-Square is largely judgmental, but it is enhanced if meaningful patterns in the configuration can be observed.

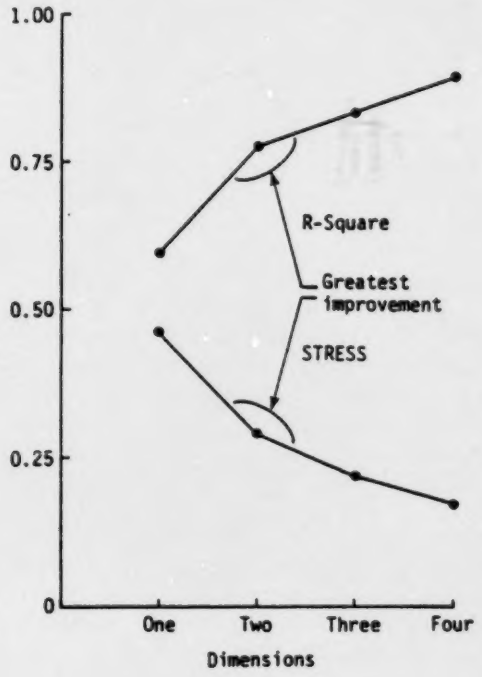

Fig. 1. STRESS and R-Square values for 60 network libraries

At this stage of analysis, no meaningful patterns in the third- and fourth-dimensional configurations have been observed. Two dimensions do show meaningful structure, however.

Figure 2 is an interpretation of the two-dimensional configuration. Symbols for each library indicate both its network and its type (public or academic).

Additionally, the x- and y-axes in the printout have been rotated approximately 10° in the interpreted version. This rotation has no effect on the configuration, but it does help the interpretation. Considered as a map, the configuration represents the location of each institution in relation to every other. The nearer each institution is to another on the map, the more similar their library collections. Conversely, the farther one institution is from another, the more dissimilar the pair is. If a library is by itself (e.g., near the perimeter), its collection is likely to be unique relative to the other collections. Of course, distances are mathematical, not geographic.

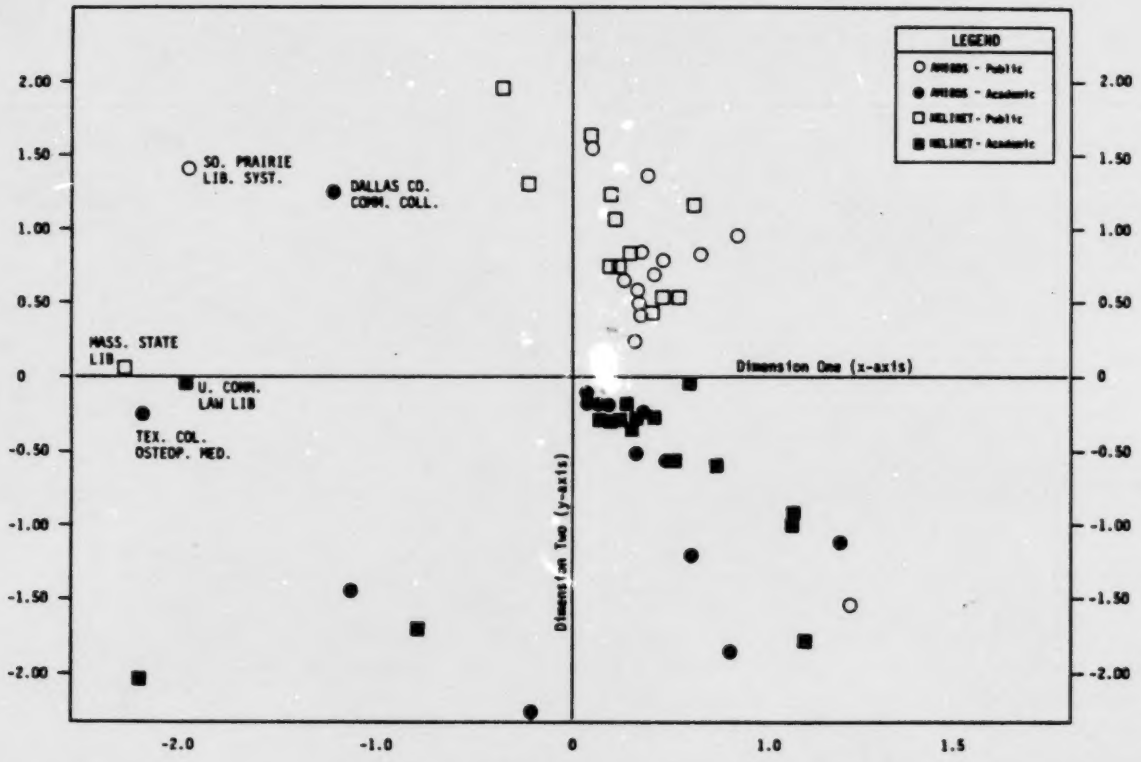

Fig. 2. Two-dimensional configuration of 60 libraries in two networks (see APPENDIX A)

Given that the map is computed and not intellectually or subjectively composed, the structure is quite dramatic. With very few exceptions, public libraries uniformly fall above, and academic libraries below, the horizontal axis. Whether in New England or the Southwest, public libraries are more similar to each other than to academic libraries in their respective geographic areas. Likewise, academic libraries are more similar to each other than they are to public libraries. Neither geographical location nor network affiliation appears to affect these similarities.

That public and academic libraries form separate clusters is not unexpected: this should be true because of their different collections and patronage. Other clusters, however, are not so obvious. Many of the larger academic libraries cluster near the center of the map. Public libraries form a large cluster in the upper-right quadrant. Certain specialized libraries (law and medical in particular) fall far to the left of the vertical axis, and a few seminary libraries were so distant from all other libraries that they could not be plotted on the map. (Though remotely dissimilar libraries might rightfully be included in a configuration--their dissimilarity merely indicates fewer holdings in

common with other libraries in the group--their exclusion would loosen the dense clustering near the center and might help to identify other patterns. One exclusion criterion is to sum the element values--total number of shared holdings--in the row/column for each library. If the total is less than the number of elements in that row/column, the library could be excluded.)

What the dimensions represent is not clear in this early stage of analysis. The x-axis (Dimension One) may be correlated with the length of time a library has been cataloging its holdings in the OCLC Online Union Catalog. That is, two libraries that began their current cataloging at the same time would have more books with the same imprint year than two libraries that began online cataloging at different times, unless one has done considerable retrospective conversion. On the y-axis, a library on the upper end may serve a broader patronage (the general public), while those on the lower end may serve a more specialized one (academics). Further analysis and hypothesis testing needs to be done before these dimensions are satisfactorily interpreted.

These results have also been reported in a separate paper (McGrath 1980a).

4.2 50 Academic Libraries

Results for the academic library sample are shown in table 4. STRESS and R-Square values are comparable to those for the 60 network libraries, showing improvement with each higher dimensional solution. Like the network sample, the sharpest improvement in STRESS occurs from the one- to the two-dimensional solution, indicated by the greater angle between dimension one and three (fig. 3). On the other hand, R-Square improvement is less sharp from dimension three to dimension four, suggesting that three dimensions should be explained.

TABLE 4

Configurations of STRESS and R-Square Values for 50 Academic Libraries

Configuration (Dimension)	STRESS	R-Square	Comment
Four	0.19	0.82	Excellent, but probably not interpretable
Three	0.25	0.76	Good
Two	0.34	0.66	Adequate and interpretable
One	0.49	0.55	Poor and inadequate

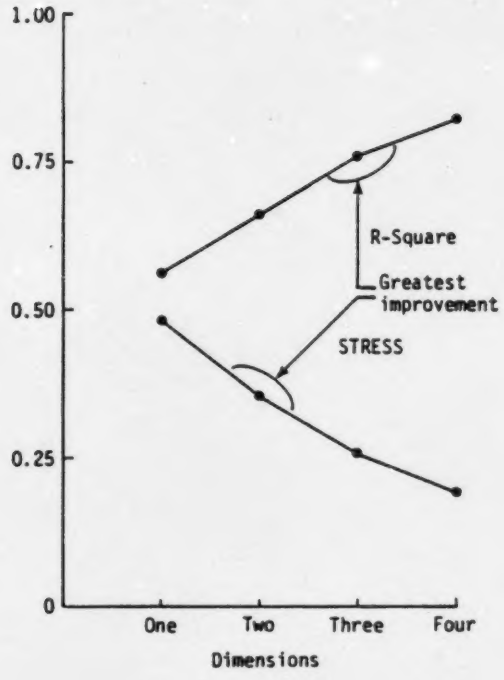

Fig. 3. STRESS and R-Square values for 50 academic libraries

The three-dimensional solution was inspected, but no interpretable pattern was observed. This is not to say, however, that one cannot be found. Additional analysis must be done if the third dimension is to be interpreted. Two dimensions, on the other hand, are quite interpretable.

Earlier overlap studies sometimes used Venn diagrams to illustrate the number of books held in common by a group of libraries (see figure 4). For example, Libraries A and B have five books in common, while A, B, and C have three. The Venn diagrams are always circular; they have no other meaning than to indicate overlap.

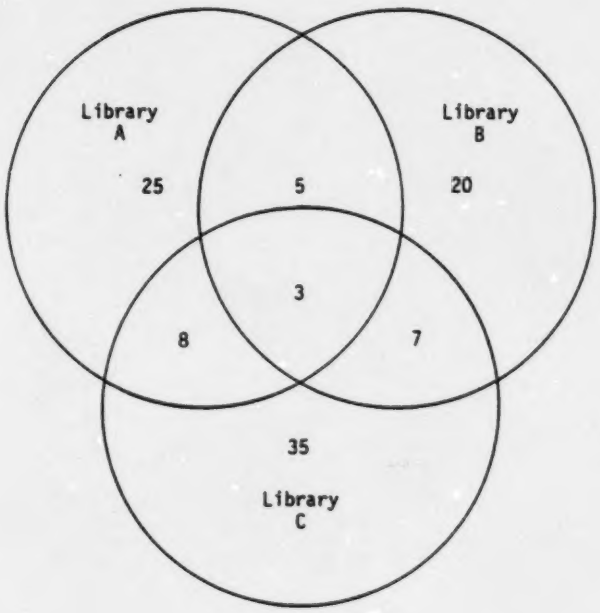

Fig. 4. Generalized Venn diagram in early overlap studies; books held in common by libraries A, B, and C

In contrast, the circles and ellipses in figure 5 (with letters replaced by symbols as indicated) are intended to encompass in the true sense. They encompass specific libraries and their locations relative to others, instead of mere numbers as in figure 4. Furthermore, they show overlap by type of library. Except for a few outliers, the perimeters show the outside range of types.

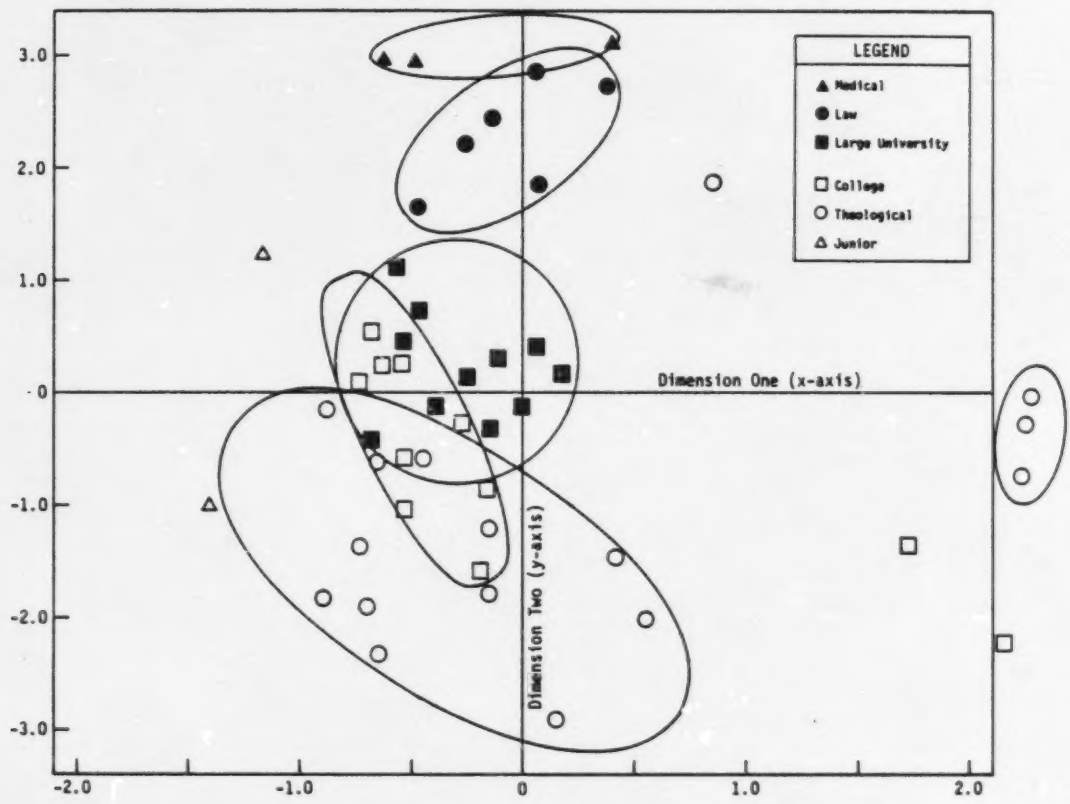

Fig. 5. Two-dimensional configuration of 50 academic libraries (see APPENDIX B)

Each library was identified as medical, law, large university, college, theological and denominational, and junior, either by the name of its respective institution or by information about its holdings published in the American Library Directory (Jacques Cattell Press 1980). These identifications have not been verified with the respective institutions.

A definite pattern or progression of type from top to bottom can be seen. Except for some intermingling near the center, libraries with solid symbols are almost completely separated from those with open symbols, with most of the solids above the x-axis and the opens below.

Three libraries form a line across the top, close together on the y-axis and slightly separated on the x-axis. Two of these are medical, and one is a community college with holdings in dental hygiene, nursing, and police science.

Law libraries form a cluster that reveals considerable homogeneity, but their proximity to medical libraries indicates something more in common with each other than with other groups.
University libraries cluster near the center, mostly above the x-axis, clearly separated from law and medical libraries.

College libraries are also clustered near the center, with some scattering below and to the left, but still pretty much within the circle of university libraries. This indicates that no clear distinction exists between the holdings of college and university libraries within this circle.

Theological and denominational libraries fall out uniformly below the x-axis, somewhat overlapping college libraries, but more spread out than the other groups, indicating considerably more diversity among themselves. Three seminary libraries, not in the main cluster, literally fall off the map to the far right. One denominational library, by itself to the upper right, might have both law and medical holdings.

Two junior (community) colleges are off to the left, somewhat distant from each other on the y-axis, but somewhat closer on the x-axis.

The meaning of this progression from the high to the low end of the y-axis (Dimension Two) is not clear. That it is a continuum of some sort is a safe conclusion, however. The dimension appears to be some sort of humanistic one, with professional institutions at one end, general and liberal arts in the center, and institutions concerned with human values or conscience at the other.

The horizontal dimension is more difficult to interpret and requires deeper study. It might represent some spread in theological or denominational emphasis or orthodoxy. It would be too simple to conclude that those on the right were orthodox and those on the left more secular.

In general, the similarities and dissimilarities occurred where one might expect. Law libraries were similar to each other, as were medical libraries, and so on. In some cases, however, there were similarities between different types--a college library close to a university library, for example. In other cases, considerable dissimilarity or diversity within types--such as the denominational group--was evident. One cannot always assume, therefore, that homogeneity of type implies similarity of collections.

As with the network results, these are rich with implications and augur many insights for deeper analysis.

Lastly, no network or geographic pattern could be observed. Geographic distance appears to have no effect on similarity or dissimilarity of collections. This conclusion was reached by visual inspection: network symbols were substituted for letter symbols. The apparent random distribution of these symbols in four quadrants of the two dimensions did not warrant further analysis (table 5).

TABLE 5

Graphic Distribution of 50 Academic Libraries by Network

Network	Quadrant			
	Upper left	Upper right	Lower left	Lower right
AMIGOS	1	1	1	1
BCR	1		1	
CAPCON	1			
FEDLINK			1	
ILLINET				1
INCOLSA				1
MIDLNET	2			
MINITEX			2	1
MLC	2		2	
NELINET			5	2
OHIONET		2	1	
PALINET			2	2
PRLC			1	
PRLC-2			1	
SOLINET	2		1	2
SUNY	4	2	2	
WESTERN	1		1	

NOTE: Distribution of the two-dimensional configuration (four quadrants) for 50 academic libraries is summarized here by network affiliation.

4.3 50 Public Libraries

Results for the 50 public libraries are shown in table 6. Other than an inspection to observe geographic or network patterns, little interpretation of this sample was undertaken. Figure 6 suggests, however, that two dimensions, like the 60-network and 50-academic samples, are adequate to explain possible relationships. Figure 7 shows three possible clusters in two dimensions, but no attempt was made to interpret them.

TABLE 6

Configurations of STRESS and R-Square Values for 50 Public Libraries

Configuration (Dimension)	STRESS	R-Square	Comment
Four	0.19	0.83	Good, but not much improvement
Three	0.24	0.79	Good
Two	0.31	0.72	Adequate
One	0.47	0.59	Poor and inadequate

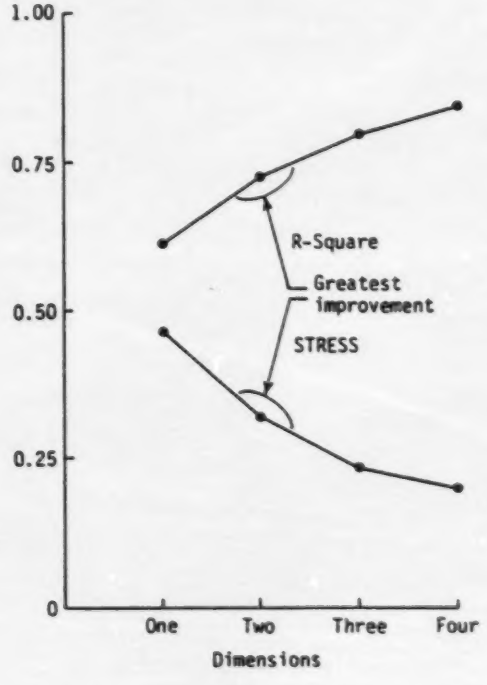

Fig. 6. STRESS and R-Square values for 50 public libraries

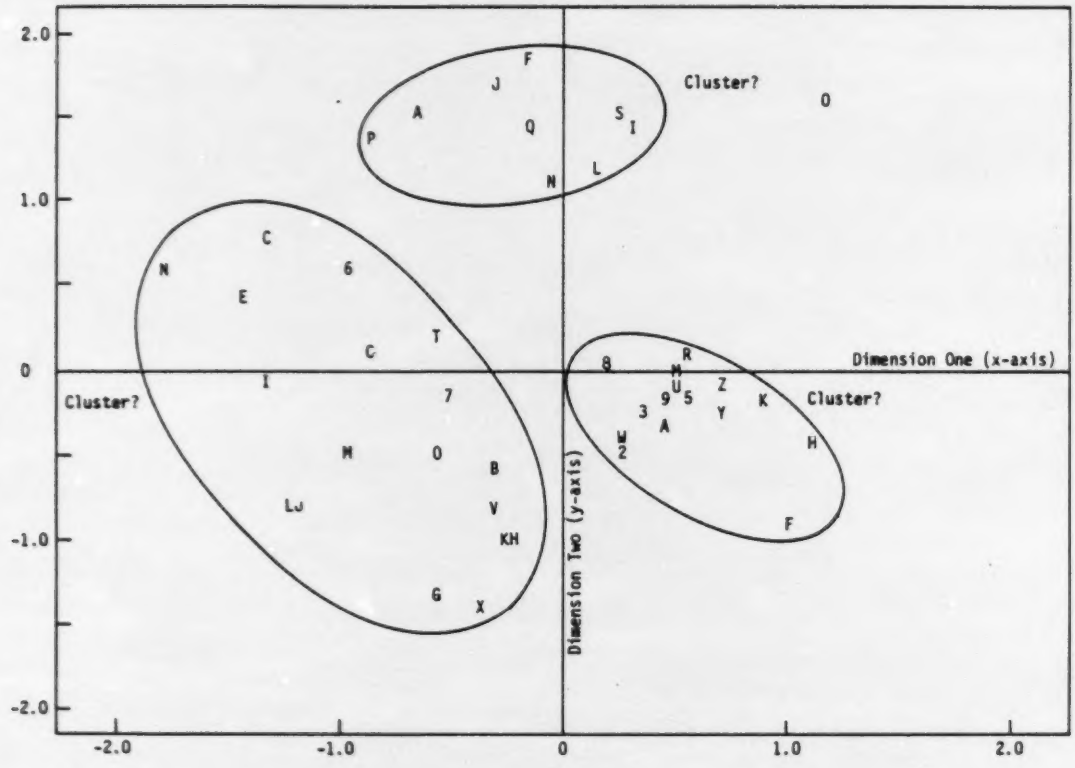

Fig. 7. Two-dimensional configuration of 50 public libraries. Possible clusters are uninterpreted.

Inspection for geographic or network pattern was done in the same manner as that for academic libraries (table 7). Again no such pattern was found.

TABLE 7

Graphic Distribution of 50 Public Libraries by Network

Network	Quadrant			
	Upper left	Upper right	Lower left	Lower right
AMIGOS		1	2	
BCR			1	
INCOLSA			1	
MLC	2		2	
NELINET	2	1	2	
OHIONET	3		2	3
PALINET	2	1		
PRLC-2	3	3		
SOLINET				5
SUNY			1	4
WILS			2	
WESTERN			1	2
UNIDENTIFIED		1		1

NOTE: Distribution of the two-dimensional configuration (four quadrants) for 50 public libraries by network affiliation.

4.4 100 Randomly Selected Libraries

If the 100 randomly selected libraries are truly representative of OCLC membership, the results should typically represent the holdings structure overall--at least as far as clusters and dimensions are concerned. Whether or not this is true will require much deeper analysis. One minor problem--that the 1% sample of holdings records was not truly random--may distort the true configuration. From a purely intuitive viewpoint, however, the number of records in the 1% sample, 58,375, should be large enough to wash out distortions between types of libraries, although some errors might occur in the distances between individual libraries.

This sample results in an average number of 11.5 holdings per cell. The range of holdings, of course, is very large--from zero (many zeroes) to several hundred. The distribution of these holdings in the 4,950-cell matrix would be an interesting study in itself.

STRESS and R-Square values for the 100 libraries appear in table 8. Values are similar to those for the other samples, and improvement occurs through four dimensions. Angular improvement (fig. 8) between dimensions one and three and two and four is about the same, however, indicating no clear preference for two, three, or four. Usually, the larger the number of objects, the larger the number of dimensions to explain the configuration. It is possible that, with this sample, the "el," or larger angle, occurs between three and five (not computed), indicating that four dimensions are needed. Considering the results of the other three samples, it is unlikely that more than three dimensions are needed.

TABLE 8

Configurations of STRESS and R-Square Values for 100 Randomly Selected Libraries

Configuration (Dimension)	STRESS	R-Square	Comment
Four	0.21	0.80	Good, but little improvement
Three	0.25	0.75	Good
Two	0.35	0.67	Adequate
One	0.53	0.54	Poor and inadequate

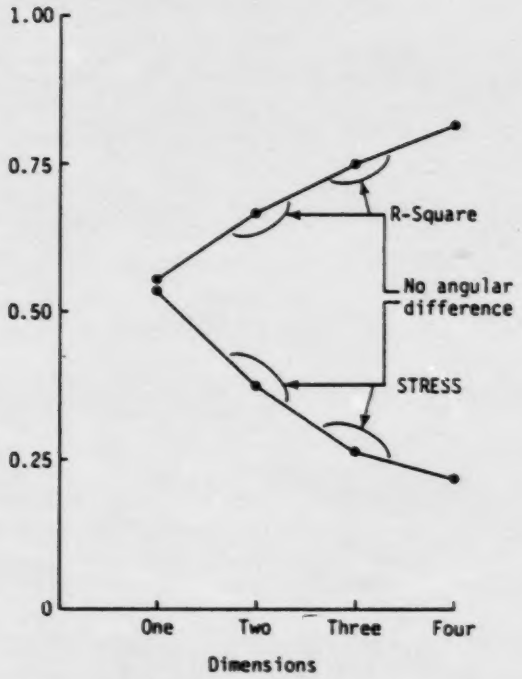

Fig. 8. STRESS and R-Square values for 100 randomly selected libraries

Inspection of the two-dimensional configuration (fig. 9) shows clustering not unlike that for the academic library sample. The x- and y-axes have been rotated about 60° clockwise to aid interpretation. The rotated axes clearly divide library types, with general academic and public libraries above the x-axis and special libraries below. Denominational and liberal arts college libraries fall to the left of the y-axis and large college, university, and public libraries to the right.

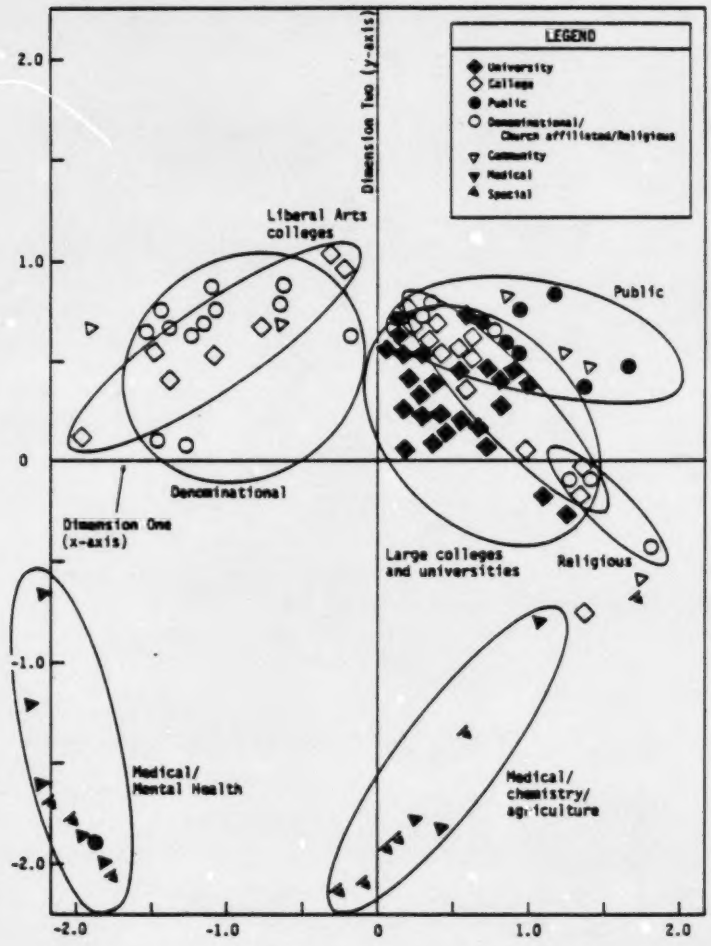

Fig. 9. Two-dimensional configuration of 100 randomly selected libraries. Clusters and overlap shown by type. (See APPENDIX D.)

Large college libraries are densely concentrated within the larger concentration of university libraries near the center of the map.

Public libraries, several of them quite large, overlap the large college libraries.

A cluster of liberal arts colleges cuts across a more diffuse cluster of denominational or church-affiliated institutions in the upper left quadrant.

In the lower right quadrant, a small cluster of religious libraries, some distance away from the denominational cluster, appears to brush against the college and university cluster, a proximity that may be better explained in higher dimensions.

In the lower right quadrant, there is an elongated cluster that includes chemistry, medical, and agriculture libraries. The library at the extreme right-hand tip of that cluster is church-affiliated, with medical and law holdings.

At the very edge of the lower-left quadrant are a number of medical, health, and other special libraries. Included in this group is a state book processing center (Florida) and several special libraries which may have health-related holdings. It should be noted that several libraries in this cluster do not appear on the computer printout since their extreme coordinates in the unrotated Dimension One could not be accommodated in the centering algorithm of the ALSCAL program. They are plotted by hand in figure 9.

Scattered throughout the configuration are the junior and community colleges which exhibit no particular pattern except where they may be included in one of the identifiable clusters.

This random sample, twice the size of the academic and public samples, inevitably introduces greater complexity of interrelationships, a complexity quite apparent in this two-dimensional configuration. Nevertheless, understandable patterns are visible and may be even more visible in the three-dimensional configuration.

5.0 IMPLICATIONS, APPLICATIONS, AND CONCLUSIONS

5.1 Network Libraries

The two-dimensional configuration for the 60-network sample suggests that network databases might be partitioned to take advantage of similarity patterns. Since there is a very strong tendency for public and academic library holdings to dichotomize, it might make sense to store them in separate files of the database, reducing retrieval time for both cataloging and interlibrary loan. In general, the more distant the library types are from each other in a configuration, the stronger the argument for separating their holdings online. There is little rationale for storing together two dissimilar types of files (e.g., bibliographic records and airline schedules).

Record storage by subject area (e.g., fiction, science/technology, etc.) is another possibility. A distinction should be made, of course, between what is accessed and by whom (type of library). A controlled experiment to study these questions would be most informative.

5.2 Academic Libraries

The two-dimensional configuration for the 50-academic-library sample has strong implications for cooperative collection development. These implications have also been reported in a separate paper (McGrath 1982).

An objective of resource sharing ought to be some ideal balance between similarity and dissimilarity. Greater similarity would help libraries find titles they wished to catalog; greater dissimilarity would help libraries borrow from others. Cooperative collection development (CCD) should help achieve that balance.

The particular patterns discussed above suggest some possibilities for CCD. The above observations are very general and idealized. Since this study contains virtually no information about individual libraries, other than name similarity to other libraries, it cannot pretend to make recommendations for a particular library. CCD also has to be done within the context of individual library policy.

Homogeneous types of libraries, such as medical and law libraries, would have much to gain by pooling their budgets to build exhaustive collections within the specific subject areas that they jointly serve. Each library would not need to have a copy of every book or journal, so long as it could be found somewhere in the group. This policy would result in greater diversity of title holdings and greater strength within subject areas.

On the other hand, university libraries, which appear to have more similarity as a group, should diversify more. These universities, with large acquisition budgets, all seem to be acquiring the same books and journals. They could achieve greater strength by acquiring materials in as many different subject areas as possible. Instead of each library being all things to all people, it should specifically profile its institutional program by subject and reexamine all blanket order plans (in which it receives everything from a publisher, regardless of the subject). This suggestion does not contradict the orthodox policy of exhaustive collection building for Ph.D. programs, since drastic modification of that policy might be difficult to achieve at many universities. It does imply that large libraries should make much closer analyses of their audiences, agreeing to exhaustive collection development within a variety of specific subject areas.

Law and medical libraries, because of some apparent similarities between the two types, might want to study the nature of their similarity to determine more precisely their common interest (e.g., the legal aspects of medicine, medical regulations, medical suits, forensics, medical ethics, etc.). As a part of such a study, answers to these questions should be found: Do medical libraries acquire legal materials? Do law libraries acquire medical materials? Should medical and law libraries form a cooperative network, formal or informal?

Theological and denominational libraries, of course, have their own special interest and unique emphasis. Their diversity indicates a rich array of subject matter that is very different from other types of libraries. It is not necessary to suggest that these libraries assign areas of collection development responsibility. They are already doing it by acquiring materials specific to their denominational interest and the other subject areas they emphasize. Their collections undoubtedly emphasize liberal arts and humanistic subjects, but these areas are so rich in resources that further diversification may not be necessary.

In this random group of libraries, the most obvious characteristic is the diversity, with no common focus overall. Within this diversity, however, there is considerable overlapping of library type--primarily by college, university, and denominational--and some apparent overlap among law and medical libraries. On the other hand, overlap between other types, such as medical and denominational, is minimal.

It is not too often that a library of one type, such as a medical library, would wish to borrow from another, dissimilar type, such as a Baptist college library. A medical library would be more likely to need something held by another medical library. It would seem, therefore, that homogeneous types would have more to gain from cooperative collection development in terms of achieving greater depth. On the other hand, large general libraries would have more to gain by achieving greater diversity. These conclusions are by no means definitive, but should be regarded as topics for further exploration and debate.

5.3 Public Libraries

Implications for the public library sample have not been explored. However, the fair two dimensionality and apparent clustering implies a meaningful structure, and hence the implications for storage allocation and retrieval and, perhaps, the sharing of materials and the development of interlibrary lending relationships. This structure warrants further inspection and analysis.

5.4 100 Randomly Selected Libraries

Implications for the 100-randomly-selected-library sample have not been thoroughly explored. However, the apparent clustering by type of library (public, academic, special, small college, and university), which agrees with the findings for the network and academic samples, reinforces the suggestion for both an in-depth analysis of holdings by type of library and the need to determine whether separate storage and access by type of library may reduce retrieval time.

If the entire database is truly structured by type of library, and if the present architecture of the database would permit restructuring, there might well be possible cost- and time-saving implications in these findings--at least for interlibrary lending, if not for cataloging.

Another possibility would be to create distributed databases by library type for specialized access and retrieval. Special libraries might find it advantageous to pool their holdings and to write special subject retrieval programs.

Libraries having similar materials may want to equalize the costs of interlibrary loan by sharing on a one-to-one basis or by some other formula based on mutual holdings.

5.5 Contribution of Multidimensional Scaling and Other Techniques

Some debate exists among statisticians about whether multidimensional scaling is a hypothesis-testing technique. This issue will not be discussed here, but, in general, the results of MDS in this study are both descriptive and hypothesis generating. In other words, a multidimensional configuration can be considered as a new way to describe the relationships between libraries. The configurations are in terms of clusters, and the dimensions of those configurations, with their Cartesian coordinates as values, may be regarded as new descriptive and independent variables.

For example, in addition to well-known variables, such as the number of staff members, number of subscriptions, number of books added, materials, budget, enrollment, and so on, we now have dimensions one, two, three, or four, depending on the values of STRESS and R-Square, and possibly even five and six. These new variables may then be used with standard hypothesis-testing techniques, such as analysis of variance and multiple regression, to "explain" the configuration or relationships. These new variables may be tested in any theoretical prediction equation, such as an allocation model. Additionally, they may also be important, when more

fully developed, in evaluating services to the patronage. For example, we may hypothesize that a library with a high (or low) value on Dimension One or Two is more effective than a library with a value on the other end of the dimension. It would be most interesting to test these new variables against such dependent variables as user satisfaction and annual circulation.

Though formal statistical hypothesis testing was not done in this project, it is entirely possible to form a hypothesis, to make a visual inspection of the configuration, and then to judge whether the configuration supports that hypothesis. For example, it could have been hypothesized that types of libraries will form homogeneous clusters (public libraries will cluster together, separate from academic libraries) and that geographic proximity has no relationship to holdings similarity.

This is not a difficult hypothesis to formulate. Based on our general understanding of libraries, it is what we should expect. Figure 2 graphically supports this "hypothesis." A better hypothesis would have sought to explain why types of libraries cluster, and a still better one why libraries are aligned in more than one dimension. For example, why should two libraries be close to each other on one dimension, but distant on another?

Understanding these dimensions and their interpretation is perhaps the most intriguing challenge to come out of this study. Some questions that naturally arise out of this dimensionality are: What is Dimension One, and how does it differ from Dimension Two, Three, or Four? Do the dimensions for each group of libraries differ, or are they highly correlated? In general, what are their correlates?

Answers to these and other questions would provide an important contribution to library network theory because good theory enables one to explain phenomena, while a satisfactory explanation would enable one to predict. For example, if certain characteristics of a library were known, but no prior knowledge of its holdings were, then the multidimensional explanation would enable one to predict which other libraries in the system had holdings similar to it.

Multidimensional scaling, in combination with both standard and new techniques (such as cluster analysis, tree fitting, multiple regression, and discriminant analysis), should help to answer these questions (Shepard 1980).

One interesting application of multidimensional scaling is the work being done by Henry Small (1973), Belver Griffith (Small and Griffith 1974), and others (Griffith et al. 1974), and H. D. White and Griffith (1981), in which the objects compared are either documents or researchers. Data consist of co-citations (i.e., the number of times any two documents or authors are cited together), indicating some similarity or other relationship between them. The results are configurations of documents or individuals in subject space, representing "invisible colleges" in narrow disciplines.

Another application was performed by the principal author (McGrath 1980b), in which broad disciplines were the objects compared, and the data were the

number of library books in those disciplines charged out by students enrolled in those disciplines. The results were configurations of disciplines in three dimensions.

Together with the results of this report, these studies can be described as generating three levels of information and knowledge space: the individuals in subdiscipline spaces, broad disciplines in a university space, and library collections in network spaces.

The structure of these and other spaces, once fully mapped and explained, show promise of substantial contribution to library and information theory.

6.0 FURTHER RESEARCH

First, clusters in the two-dimensional, public library configuration should be interpreted.

Second, greater effort should be made to interpret the third and fourth dimensional solutions. If no valid interpretations are found, they may simply be "noise" or error.

Third, an effort should be made to identify interpreted dimensions. One way to do this is to correlate Cartesian coordinates with other known characteristics or variables.

Fourth, a search of the holdings database might be made in order to determine the most similar and dissimilar libraries. This knowledge might be used to take better advantage of holdings for cataloging and interlibrary loan. The concept of a "most similar libraries' network" and a "most dissimilar libraries' network" might be worth exploring.

Last, the finding that geography or political region has little or no relationship to holdings in a network should be studied further. The finding is important because virtually all networks have been organized according to geography or political region.

REFERENCES

Altman, Ellen. 1972. Implications of title diversity and collection overlap for interlibrary loan among secondary schools. Library Quarterly 42(2): 177-194.

Davis, Charles, and D. Shaw. 1979. Collection overlap as a function of library size: A comparison of American and Canadian public libraries. Journal of the American Society for Information Science 30(1): 19-24.

Evans, Glyn, Mary Beilby, and Roger Gifford. 1978. Development of a responsive library acquisitions formula: Final report. Washington, D.C.: U.S. Office of Libraries and Learning Resources.

Griffith, Belver C., Henry G. Small, J.A. Stonehill, and S. Dey. 1974. The structure of scientific literature II: Toward a macro- and micro-structure for science. Science Studies 4(4): 339-365.

Jacques Cattell Press. 1980. American library directory. 33rd ed. New York and London: R.R. Bowker Co.

Kilgour, Frederick G. 1969. The economic goal of library automation. College and Research Libraries 30(4): 307-311.

Kruskal, Joseph. 1964. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika 29(1): 1-27.

Kruskal, Joseph B., and Myron Wish. 1978. Multidimensional scaling. Sage university paper series on quantitative applications in the social sciences. No. 7-11. Beverly Hills and London: Sage Publications.

Leonard, Lawrence E., Joan M. Maier, and Richard M. Dougherty. 1969. Centralized book processing: A feasibility study based on Colorado academic libraries. Metuchen, NJ: Scarecrow Press.

McGrath, William E. 1980a. Multidimensional map of library similarities. In Communicating information: Proceedings of the 43rd annual ASIS meeting 17:298-300. White Plains, NY: Knowledge Industry Publications.

------. 1980b. The circulation uncertainty principle and the cosmology of collection use. In Proceedings of the fourth international conference on approval plans and collection development. Phoenix: Oryx Press.

------. 1982. Implications for cooperative collection development in a random group of academic libraries: Or, beyond overlap. Eds. Michael D. Kathman and Virgil F. Massman. In Options for the eighties: Proceedings of the second national conference of the Association of College and Research Libraries. Greenwich, CT: JAI Press.

McGrath, William E., Donald Simon, and Peter Dickinson. 1971. Predicting title multiplication (overlap) in a union catalog of sixteen Louisiana libraries using regression analysis. Available from ERIC, Arlington, VA: Doc. no. Li 004 023.

Merritt, Leroy C. 1942. The administrative, fiscal, and quantitative aspects of the regional union catalog. In Union catalogs in the United States. Chicago: American Library Assoc.

Nugent, William. 1968. Statistics of collection overlap at the libraries of the six New England state universities. Library Resources and Technical Services 12(1): 31-36.

Shepard, Roger N. 1962. The analysis of proximities: Multidimensional scaling with an unknown distance function. Psychometrika 27(3): 219-246.

------. 1980. Multidimensional scaling, tree-fitting and clustering. Science 210: 390-398.

Small, Henry G. 1973. Co-citation in the scientific literature: A measure of the relationship between two documents. Journal of the American Society for Information Science 24: 265-269.

Small, Henry G., and Belver C. Griffith. 1974. The structure of scientific literature I: Identifying and graphing specialties. Science Studies 4(2): 17-40.

Torgerson, W.S. 1958. Theory and methods of scaling. New York: John Wiley.

White, Howard D., and Belver C. Griffith. 1981. Author cocitation: A literature measure of intellectual structure. Journal of the American Society for Information Science 32(3): 163-171.

Young, Forrest W., and Rostyslaw Lewyckyj. 1979. ALSCAL-4 user's guide; A guide for users of ALSCAL-4: A nonmetric multidimensional scaling and unfolding program with several individual differences options. Carrboro, NC: Analysis and Theory Associates.

APPENDIX A

RANDOM SAMPLE LIST OF 60 NETWORK LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	TYPE	NETWORK	ALSCAL SYMBOL
		AMIGOS PUBLIC				
579	AZT	Tucson Pub Libr	Arizona	Public	AMIGOS	A
246	IGA	Dallas Pub Libr	Texas	Public	AMIGOS	B
229	IJC	Irving Pub Libr	Texas	Public	AMIGOS	C
1336	OKE	Metropolitan Libr System	Okla	Public	AMIGOS	D
666	OKP	S Prairie Libr System	Okla	Public	AMIGOS	E
1319	TUL	Tulsa City-County Libr	Okla	Public	AMIGOS	F
609	TXB	Abilene Pub Libr	Texas	Public	AMIGOS	G
618	TXG	Austin Pub Libr	Texas	Public	AMIGOS	H
585	TXL	Lubbock City-County Libr	Texas	Public	AMIGOS	I
850	TXN	Houston Pub Libr	Texas	Public	AMIGOS	J
593	TXP	El Paso Pub Libr	Texas	Public	AMIGOS	K
1331	YCC	Yuma City-County Libr	Arizona	Public	AMIGOS	L
228	IFA	Fort Worth Pub Libr	Texas	Public	AMIGOS	M
580	TXW	Waco-McLennan County Libr	Texas	Public	AMIGOS	N
668	AKE	Arkansas Libr Commission	Ark	Public	AMIGOS	O
		NELINET PUBLIC				
211	BYN	Bangor Pub Libr	Maine	Public	NELINET	P
126	FEM	Ferguson Pub Libr	Conn	Public	NELINET	Q
205	HPL	Hartford Pub Libr	Conn	Public	NELINET	R
162	NWK	Norwalk Pub Libr	Conn	Public	NELINET	S
157	PEM	Perrot Memorial Libr	Conn	Public	NELINET	T
201	PPN	Portland Pub Libr	Maine	Public	NELINET	U
1254	QUI	Thomas Crane Pub Libr	Mass	Public	NELINET	V
986	SML	S Maine Libr District	Maine	Public	NELINET	W
125	SSA	Stamford Pub Libr	Conn	Public	NELINET	X
374	WZW	Worcester Pub Libr	Mass	Public	NELINET	Y
206	CZL	Connecticut State Libr	Conn	Public	NELINET	Z
189	MEA	Maine State Libr	Maine	Public	NELINET	1
235	RDS	RI Dept of Libr Serv	RI	Public	NELINET	2
1789	GRN	Greenwich Libr	Conn	Public	NELINET	3
1714	MAS	Massachusetts State Libr	Mass	Public	NELINET	4
		AMIGOS ACADEMIC				
611	AKH	Henderson State Univ	Arizona	Academic	AMIGOS	5
981	AKK	John Brown Univ	Arizona	Academic	AMIGOS	6
197	IDA	Dallas Baptist Col	Texas	Academic	AMIGOS	7
209	IRU	New Mexico State Univ	New Mex	Academic	AMIGOS	8
591	TXH	Univ of Houston	Texas	Academic	AMIGOS	9
984	NMA	Univ of Albuquerque	New Mex	Academic	AMIGOS	A
1264	OKC	Cameron Univ	Okla	Academic	AMIGOS	B
815	OKG	Phillips Univ Grad Sem Libr	Okla	Academic	AMIGOS	C

APPENDIX A (cont.)

RANDOM SAMPLE LIST OF 60 NETWORK LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	TYPE	NETWORK	ALSCAL SYMBOL
		AMIGOS ACADEMIC				
677	OKS	Oklahoma State Univ	Okla	Academic	AMIGOS	D
1279	TDJ	Dallas County Commun Col	Texas	Academic	AMIGOS	E
1281	TOM	Texas Col of Osteopathic Med	Texas	Academic	AMIGOS	F
635	TX1	Southwest Texas State Univ	Texas	Academic	AMIGOS	G
606	TXR	Lamar Univ	Texas	Academic	AMIGOS	H
583	AZS	Arizona State Univ	Arizona	Academic	AMIGOS	I
613	TXA	Texas A & M Univ	Texas	Academic	AMIGOS	J
		NELINET ACADEMIC				
324	AZM	Assumption Col	Mass	Academic	NELINET	K
127	BHA	Harvard Divinity Sch	Mass	Academic	NELINET	L
536	BTS	Bates Col	Maine	Academic	NELINET	M
76	DRB	Dartmouth Col	NH	Academic	NELINET	N
1041	HCL	Husson Col	Maine	Academic	NELINET	O
1296	MBB	Brandeis Univ	Mass	Academic	NELINET	P
215	MYG	Massachusetts Inst of Tech	Mass	Academic	NELINET	Q
139	PGP	Univ of S Maine at Portland	Maine	Academic	NELINET	R
94	RCM	Rhode Island Col	RI	Academic	NELINET	S
92	RIU	Univ of Rhode Island	RI	Academic	NELINET	T
1169	STO	Stonehill Col	Mass	Academic	NELINET	U
1276	UCL	Univ of Conn Law Libr	Conn	Academic	NELINET	V
326	VTJ	Johnson State Univ	Vermont	Academic	NELINET	W
183	SNN	Smith Col	Mass	Academic	NELINET	X
99	YUS	Yale Univ	Conn	Academic	NELINET	Y

APPENDIX B

RANDOM SAMPLE LIST OF 50 ACADEMIC LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	NETWORK	ALSCAL SYMBOL
8	CIN	Univ of Cincinnati	Ohio	OHIONET	A
29	AKR	Univ of Akron	Ohio	OHIONET	B
67	FMC	Florida Mem Col	Florida	SOLINET	C
105	FAU	Fairfield Univ	Conn	NELINET	D
123	BXM	Boston Col	Mass	NELINET	E
143	SJD	St Joseph's Col	Pa	PALINET	F
163	HRM	Univ of Hartford	Conn	NELINET	G
175	CKM	Clark Univ	Mass	NELINET	H
220	WUM	Washington Univ Med Sch	Mo	MIDLNET	I
230	PDC	Commun Col of Philadelphia	Pa	PALINET	J
248	RSC	St Charles Seminary	Pa	PALINET	K
274	MKN	Northeast Missouri State	Mo	MIDLNET	L
301	ZRS	Russell Sage Col	NY	SUNY	M
304	GHA	Georgia Southwestern	Georgia	SOLINET	N
308	MQM	Univ of New Mexico Health Science Libr	New Mex	AMIGOS	O
324	AZM	Assumption Col	Mass	NELINET	P
356	WFN	Westminster Col	Pa	PRLC	Q
362	PZI	Indiana Univ of Pa	Pa	PRLC	R
366	ZDG	Corning Commun Col	NY	SUNY	S
404	GQG	Gallaudet Col	Wash DC	FEDLINK	T
418	VOC	Onondaga Commun Col	NY	SUNY	U
459	VDB	Brooklyn Col	NY	SUNY	V
523	GCL	Columbia Theological	Georgia	SOLINET	W
542	FUB	Univ of Florida Law Sch	Florida	SOLINET	X
548	EXH	Hope Col	Mich	MLC	Y
551	BTI	Boston Theological	Mass	NELINET	Z
554	EZC	Central Michigan Univ	Mich	MLC	1
580	UTL	Univ of Toledo Law Libr	Ohio	OHIONET	2
583	AZS	Arizona State Univ	Arizona	AMIGOS	3
699	VVJ	John Jay Col of Crim	NY	SUNY	4
743	ACL	Ambassador Col	Calif	WESTERN	5
752	BAL	Univ of Baltimore	Md	CAPCON	6
819	EYD	Univ of Mich-Dearborn	Mich	MLC	7
863	COF	Colorado State Univ	Colo	BCR	8
902	EZF	Ferris State Col	Mich	MLC	9
953	IBN	Blackburn Col	Ill	ILLINET	A
965	SDY	Mount Marty Col	SDAK	BCR	B

APPENDIX B (cont.)

RANDOM SAMPLE LIST OF 50 ACADEMIC LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	NETWORK	ALSCAL SYMBOL
1003	MNI	Winona State Univ	Minn	MINITEX	C
1011	MUS	Univ of S Mississippi	Miss	SOLINET	D
1033	MBS	Evan Lutheran Synod Sem	Minn	MINITEX	E
1062	CLV	La Verne Univ	Calif	WESTERN	F
1083	VZH	Hartwick Col	NY	SUNY	G
1094	LYC	Lycoming Col	Pa	PALINET	H
1157	OKR	Oklahoma Baptist Univ	Okla	AMIGOS	I
1158	MNN	Carleton Col	Minn	MINITEX	J
1215	IVU	Valparaiso Univ	Ind	INCOLSA	K
1216	VZP	Pace Univ	NY	SUNY	L
1217	AHS	Harding Univ (Col)	Ark	AMIGOS	M
1226	EMC	Emmanuel Col	Mass	NELINET	N
1268	VZT	St Joseph's Col	NY	SUNY	O

APPENDIX C

RANDOM SAMPLE LIST OF 50 PUBLIC LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	TYPE	NETWORK	ALSCAL SYMBOL
122	DMM	Dayton Montgomery Pub Libr	Ohio	Public	OHIONET	A
125	SSA	Stratford Pub Libr	Conn	Public	NELINET	B
140	GEC	Geauga County Pub Libr	Ohio	County	OHIONET	C
146	SHP	Shaker Hts Pub Libr	Ohio	Public	OHIONET	D
155	UAP	Upper Arlington Pub Libr	Ohio	Public	OHIONET	E
157	PEM	Perrot Memorial Libr	Conn	Public	NELINET	F
172	LIM	Lima Pub Libr	Ohio	Public	OHIONET	G
201	PPN	Portland Pub Libr	Maine	Public	NELINET	H
211	BYN	Bangor Pub Libr	Maine	Public	NELINET	I
234	BVP	Burton Pub Libr	Ohio	Public	OHIONET	J
280	SXP	Sunnyvale Pub Libr	Calif	Public	WESTERN	K
310	PBL	Bethlehem Pub Libr	Pa	Public	PALINET	L
314	ZTM	Mid-York Libr System	NY	System	SUNY	M
335	LKC	Lancaster Pub Libr	Pa	Public	PALINET	N
341	MPL	Monessen Pub Libr	Pa	Public	PRLC	O
347	WWC	Citizen's Libr	Pa	Public	PRLC	P
351	QNC	New Castle Pub Libr	Pa	Public	PRLC	Q
352	EPL	Erie County Libr System	Pa	System	PRLC	R
354	WDL	Warren Libr Association	Pa	Assoc	PRLC	S
355	AOP	Altoona Pub Libr	Pa	Public	PRLC	T
364	ORL	Orlando Pub Libr	Florida	Public	SOLINET	U
374	WZW	Worcester Pub Libr	Mass	Public	NELINET	V
394	YQR	Rochester Pub Libr	NY	Public	SUNY	W
410	ZNM	Nioga Pub Libr	NY	Public	SUNY	X
414	DZM	Miami-Dade Pub Libr	Florida	Public	SOLINET	Y
417	JPL	Jacksonville Pub Libr	Florida	Public	SOLINET	Z
449	LNC	New Orleans Pub Libr	La	Public	SOLINET	1
520	RTN	North County Libr System	NY	System	SUNY	2
531	PFO	Pomona Pub Libr	Calif	Public	WESTERN	3
558	VVA	S Adirondack Libr System	NY	System	SUNY	4
560	CLE	Cleveland Pub Libr	Ohio	Public	OHIONET	5
576	EAP	Easton Area Pub Libr	Pa	Public	PALINET	6
609	TXB	Abilene Pub Libr	Texas	Public	AMIGOS	7
618	TXG	Austin Pub Libr	Texas	Public	AMIGOS	8
620	OCO	Columbus Pub Libr	Ohio	Public	OHIONET	9
664	WPL	Warren Pub Libr	Ohio	Public	OHIONET	A
749	ECU	Euclid Pub Libr	Ohio	Public	OHIONET	B
753	MCD	Mary Cheney Libr	Conn	Public	NELINET	C
790	CPP	Pasadena Pub Libr	Calif	Public	WESTERN	D
844	EXE	Kent County Libr System	Minn	System	MLC	E
847	ASL	Alabama Pub Libr Service	Alabama	Service	SOLINET	F

APPENDIX C (cont.)

RANDOM SAMPLE LIST OF 50 PUBLIC LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	TYPE	NETWORK	ALSCAL SYMBOL
850	TXN	Houston Pub Libr	Texas	Public	AMIGOS	G
878	EYP	Detroit Pub Libr	Mich	Public	MLC	H
970	UUC	Salt Lake County Libr	Utah	County	BCR	I
1009	GZF	Eau Claire Pub Libr	Wis	Public	WILS	J
1028	GZD	Milwaukee Pub Libr	Wis	Public	WILS	K
1029	IGP	Gary Pub Libr	Indiana	Public	INCOLSA	L
1148	EYH	Huron Valley Libr System	Mich	System	MLC	M
1156	EYS	St Clair County Libr	Mich	Public	MLC	N
1254	QUI	Thomas Crane Pub Libr	Mass	Public	NELINET	O

APPENDIX D

RANDOM SAMPLE LIST OF 100 LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	TYPE	NETWORK	SYMBOL
8	CIN	Univ of Cincinnati	Ohio	U	OHIONET	A
11	DNU	Denison Univ	Ohio	U	OHIONET	B
22	OSU	Ohio State Univ	Ohio	U	OHIONET	C
63	BIS	Bishop Col	Texas	D	AMIGOS	D
65	CFC	Claflin Col	SC	D	CCLC	E
91	CTL	Connecticut Col	Conn	C	NELINET	F
94	RCM	Rhode Island Col	RI	C	NELINET	G
105	FAU	Fairfield Univ	Conn	D	NELINET	H
115	BUF	SUNY at Buffalo	NY	U	SUNY	I
140	GEC	Geauga County Pub Libr	Ohio	P	OHIONET	J
151	HTC	Houston-Tillotson Col	Texas	D	CCLC	K
157	PEM	Perrot Memorial Libr	Conn	P	NELINET	L
161	NLM	National Libr of Med	Md	M	FEDLINK	M
163	HRM	Univ of Hartford	Conn	U	NELINET	N
197	IDA	Dallas Baptist Col	Texas	D	AMIGOS	O
203	MTH	Mt Holyoke Col	Mass	C	NELINET	P
204	TYC	Trinity Col	Conn	C	NELINET	Q
205	HPL	Hartford Pub Libr	Conn	P	NELINET	R
209	IRU	New Mexico State Univ	New Mex	U	AMIGOS	S
264	ZBM	SUNY at Oneonta	NY	U	SUNY	T
298	GUA	Univ of Georgia	Georgia	U	SOLINET	U
305	ZVM	Mohawk Valley Commun	NY	J	SUNY	V
317	LQS	Lock Haven State Col	Pa	C	PALINET	W
327	KRS	Kearney State Col	Nebr	C	NEBASE	X
330	AQM	Amer Antiquarian Soc	Mass	S	NELINET	Y
331	RPL	Reading Pub Libr	Pa	P	PALINET	Z
332	ZOW	SUNY at Old Westbury	NY	C	SUNY	1
353	AVL	Allegheny Col	Pa	C	PRLC	2
358	EFM	Palm Beach Junior Col	Florida	J	SOLINET	3
365	ZIH	Hofstra Univ	NY	U	SUNY	4
413	SUC	Univ of S Carolina	SC	U	SOLINET	5
418	VOC	Onondaga Commun Col	NY	J	SUNY	6
423	VWM	Col of William and Mary	Va	C	BCR	7
433	UPM	Pennsylvania State Univ	Pa	U	PRLC	8
448	GOM	Macon Jr Col	Georgia	J	SOLINET	9
469	OMC	Mayo Clinic	Minn	M	INDPNDNT	A
479	ETS	E Stroudsburg State Col	Pa	C	PALINET	B
493	VYL	Lehman Col	NY	C	SUNY	C
503	KUM	Univ Kentucky Med Libr	Ky	M	SOLINET	D
508	GPM	Georgia Southern Col	Georgia	C	SOLINET	E

APPENDIX D (cont.)

RANDOM SAMPLE LIST OF 100 LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	TYPE	NETWORK	ALSCAL SYMBOL
522	FBN	State Libr of Fla Bur of Book Processing	Florida	P	SOLINET	F
569	AGN	USDA-N Research Ctr	Ill	S	FEDLINK	G
589	AZU	Univ of Arizona	Arizona	U	AMIGOS	H
626	TXT	Texas Southern Univ	Texas	U	AMIGOS	I
629	GZT	Univ of Wis, Whitewater	Wis	U	WLC	J
630	AKU	Univ of Ark, Little Rock	Ark	U	AMIGOS	K
658	VVV	Utica Col	NY	C	SUNY	L
659	OKD	Oklahoma Dept of Libr	Okla	P	AMIGOS	M
694	VOD	Old Dominion Univ	Va	U	SOLINET	N
703	TUC	Univ of Tenn-Chattanooga	Tenn	U	SOLINET	O
726	CSC	California State Col	Pa	C	PRLC	P
747	CIT	Calif Inst of Tech	Calif	U	WESTERN	Q
749	ECU	Euclid Pub Libr	Ohio	P	OHIONET	R
773	MJC	Metro Commun Col	Mo	J	INDPNDNT	S
774	FDU	Fairleigh Dickinson Univ	NJ	U	PALINET	T
782	MWS	Missouri State Col	Mo	C	INDPNDNT	U
787	MVC	Maryville Col	Mo	D	MIDLNET	V
801	DOD	Dumbarton Oaks Research	Wash, DC	S	CAPCON	W
819	EYD	Univ of Mich, Dearborn	Mich	U	MLC	X
834	CAS	Chem Abstracts Service	Ohio	S	OHIONET	Y
855	MEB	Bangor Mental Health Inst	Maine	M	NELINET	Z
863	COF	Colorado State Univ	Colo	U	NCR	1
876	MST	St Cloud State Univ	Minn	U	MINITEX	2
877	GZO	Univ of Wis, Oshkosh	Wis	U	WLC	3
884	FBC	Barry Col	Florida	D	SOLINET	4
885	GAP	Atlanta Pub Libr	Georgia	P	SOLINET	5
886	CCP	Carlsbad City Libr	Calif	P	WESTERN	6
889	ISS	St Meinrad Col	Indiana	D	INCOLSA	7
891	IES	Eli Lilly & Co	Indiana	S	INCOLSA	8
902	EZF	Ferris State Col	Mich	C	MLC	9
921	PIZ	Youth Devel Ctr-Waynesburg	Pa	S	PALINET	A
925	PIW	Woodhaven Center	Pa	M	PALINET	B
941	PIC	Craig House Technoma	Pa	S	PALINET	C
962	IOO	Northwestern College	Iowa	D	BCR	D
991	TWH	Houston Baptist Univ	Texas	D	AMIGOS	E
996	TWC	Texas Wesleyan Col	Texas	C	AMIGOS	F
999	IOL	Wahlert Mem Libr-Loras College	Iowa	D	BCR	G
1054	MMU	Univ of Missouri-Columbia Med Libr	Mo	M	MIDLNET	H
1085	UUW	Westminster Col	Utah	D	BCR	I
1090	MOI	William Jewell Col	Mo	D	MIDLNET	J
1103	VZN	Col of New Rochelle	NY	C	SUNY	K
1142	ICN	North Central Col	Ill	C	ILLINET	L
1152	EYY	Mercy Col of Detroit	Mich	M/L	MLC	M

APPENDIX D (cont.)

RANDOM SAMPLE LIST OF 100 LIBRARIES

INSTITUTION NUMBER	OCLC SYMBOL	INSTITUTION	LOCATION	TYPE	NETWORK	ALSCAL SYMBOL
1170	ECL	Emerson Col	Mass	C	NELINET	N
1191	UCP	Univ of Conn Health Libr Processing Ctr	Conn	M	NELINET	O
1195	MOE	Evangel Col Libr	Mo	D	BCR	P
1198	MOO	School of the Ozarks	Mo	C	BCR	Q
1200	OPU	Pacific Univ	Oregon	C	WESTERN	R
1207	CSM	S Calif Col	Calif	D	WESTERN	S
1221	MNP	Univ of Minn, St Paul	Minn	S	MINITEX	T
1231	PSP	Swarthmore Col Peace Collection	Pa	S	PALINET	U
1235	PRW	World Univ	PR	S	CCLC	V
1238	IDB	Illinois Central Col	Ill	J	ILLINET	W
1262	FFC	Flagler Col Libr	Florida	C	SOLINET	X
1269	CBC	Biola Col	Calif	D	WESTERN	Y
1270	WVA	Alderson-Broaddus	W Va	D	PRLC	Z
1288	NJG	Glassboro State Col	NJ	C	PALINET	1
1296	MBB	Brandeis Univ	Mass	U	NELINET	2
1298	AUS	Overseas Telecom Corp		S	INDPNDNT	3
1299	THD	Univ of Houston, Downtown	Texas	U	AMIGOS	4

Institution type:

M Medical
L Law
U University
C College
D Denominational/Church
J Junior/Community
P Public
S Special